

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/135434>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Probabilistic Numerical Methods for PDE-constrained Bayesian Inverse Problems

Jon Cockayne* Chris Oates[†] Tim Sullivan[‡]
Mark Girolami[§]

January 17, 2017

This paper develops meshless methods for probabilistically describing discretisation error in the numerical solution of partial differential equations. This construction enables the solution of Bayesian inverse problems while accounting for the impact of the discretisation of the forward problem. In particular, this drives statistical inferences to be more conservative in the presence of significant solver error. Theoretical results are presented describing rates of convergence for the posteriors in both the forward and inverse problems. This method is tested on a challenging inverse problem with a nonlinear forward model.

1 Introduction

Partial differential equations (PDEs) are challenging problems which often have no analytical solution and must be solved numerically. In the style of Probabilistic Numerics (PN) [7], in this work we describe methods for probabilistically modelling the uncertainty in the true solution arising from the numerical approximation. This uncertainty can be thought of as arising from finite computation, as formalised in the Information Complexity literature [11]; in solving a problem numerically, we are forced to discretise some aspect of it. In the present work we model this uncertainty as arising from taking a finite number of evaluations of the forcing terms of the system of PDEs.

One of the core principles of probabilistic numerics is that, in complex procedures in which multiple numerical approximations must be composed to produce a final result, the uncertainty from each procedure can combine in a nontrivial way which can lead to incorrect inferences. The example we take here is that of PDE constrained Bayesian

*University of Warwick, j.cockayne@warwick.ac.uk

[†]University of Technology Sydney, chris.oates@ncl.ac.uk

[‡]Free University of Berlin and Zuse Institute Berlin, sullivan@zib.de

[§]Imperial College London and Alan Turing Institute, m.girolami@imperial.ac.uk

inverse problems, in which we wish to estimate parameters of a PDE model in a Bayesian framework, based on observations of a system which is believed to be described by the underlying PDE. In such problems it has been shown that employing an inaccurate PDE solver in the sampling can lead to incorrect inferences in the inverse problem [3].

There has been recent interest in construction of probabilistic solvers for PDEs. Work by [3] constructs a nonparametric posterior distribution for ODEs and PDEs by injecting noise into standard numerical solvers in such a way as to maintain the convergence properties of these solvers. In [8], the authors discuss a meshless method which is similar to the method discussed herein by modelling the forcing of the PDE. This is developed in [9], which discusses a methodology for probabilistic solution of PDEs by an hierarchical game-theoretic argument. These latter two approaches do not examine application to inverse problems, however.

Work from [1] discusses the interpretation of symmetric collocation as the mean function of a Gaussian process prior after conditioning on observed values of the forcing, but applies this methodology predominantly to stochastic differential equations.

1.1 Structure of the Paper

We begin by introducing the concept of a probabilistic meshless method and giving some theoretical results related to it. We then show how the posterior measure over the forward solution of the PDE can be propagated to the posterior measure over parameters in a Bayesian inverse problem. Finally we present some numerical results for a challenging nonlinear inverse problem given by the steady-state Allen–Cahn equations.

Proofs for the presented theorems are omitted, and can be found in [2].

2 The Probabilistic Meshless Method

We now introduce the concept of a probabilistic meshless method (PMM). Consider an open, bounded subset D of \mathbb{R}^d with Lipschitz boundary ∂D . We seek a solution $u \in H(D)$, some Hilbert space of functions defined over D , of the following system of operator equations

$$\begin{aligned}\mathcal{A}u(\mathbf{x}) &= g(\mathbf{x}) & \mathbf{x} \in D \\ \mathcal{B}u(\mathbf{x}) &= b(\mathbf{x}) & \mathbf{x} \in \partial D.\end{aligned}\tag{1}$$

Here $\mathcal{A} : H(D) \rightarrow H_{\mathcal{A}}(D)$ and $\mathcal{B} : H(D) \rightarrow H_{\mathcal{B}}(D)$ with $g \in H_{\mathcal{A}}(D)$ and $b \in H_{\mathcal{B}}(D)$. \mathcal{A} is associated with a partial differential operator and \mathcal{B} is associated with the boundary conditions of the system. For notational simplicity we restrict attention to systems of two operators, however the methods discussed can be generalised to an arbitrary number of operator equations.

We proceed in a Bayesian setting by placing a prior measure Π_u on u , and determining its posterior distribution based on a finite number of observations of the system given in Eq. 1. In this work we focus on the most direct observations of said system; namely, we choose sets of design points $\{\mathbf{x}_{i,\mathcal{A}}\} = X_0^{\mathcal{A}} \subset D$, $\{\mathbf{x}_{j,\mathcal{B}}\} = X_0^{\mathcal{B}} \subset \partial D$ for $i = 1, \dots, m_{\mathcal{A}}$,

$j = 1, \dots, m_{\mathcal{B}}$. We then evaluate the right-hand-side corresponding to each of the operators in the system at these points; $\mathbf{g} = [g(\mathbf{x}_{i,\mathcal{A}})]$, $\mathbf{b} = [b(\mathbf{x}_{i,\mathcal{B}})]$.

It remains to specify our prior distribution. Here we choose a Gaussian process prior $\Pi_u = \mathcal{GP}(m, k)$. Recall that a Gaussian Process is characterised by its mean function m and its covariance function k , and the property that, if $u \sim \mathcal{GP}(m, k)$ then for any set of points $\{\mathbf{x}_i\} \subset \mathbb{R}^d$, $i = 1, \dots, n$

$$\begin{aligned} u(X) &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \\ [\boldsymbol{\mu}]_i &= m(\mathbf{x}_i) \\ [\Sigma]_{ij} &= k(\mathbf{x}_i, \mathbf{x}_j) \end{aligned}$$

As is common in the literature we will use a centred Gaussian process prior; $\Pi_u = \mathcal{GP}(0, k)$. Define

$$\mathcal{L} = \begin{bmatrix} \mathcal{A} \\ \mathcal{B} \end{bmatrix} \quad \bar{\mathcal{L}} = [\bar{\mathcal{A}} \quad \bar{\mathcal{B}}]$$

and furthermore for sets $X = \{x_i\}$, $i = 1, \dots, N$, $Y = \{y_j\}$, $j = 1, \dots, M$ let $K(X, Y)$ denote the Gram matrix of K applied to X and Y ; $[K(X, Y)]_{ij} = k(x_i, y_j)$. Similarly $[\mathcal{A}K(X, Y)]_{ij} = \mathcal{A}k(x_i, y_j)$, etc. Then

$$\begin{aligned} \mathcal{L}\bar{\mathcal{L}}K(X_0, X_0) &= \begin{bmatrix} \mathcal{A}\bar{\mathcal{A}}K(X_0^{\mathcal{A}}, X_0^{\mathcal{A}}) & \mathcal{A}\bar{\mathcal{B}}K(X_0^{\mathcal{A}}, X_0^{\mathcal{B}}) \\ \mathcal{B}\bar{\mathcal{A}}K(X_0^{\mathcal{B}}, X_0^{\mathcal{A}}) & \mathcal{B}\bar{\mathcal{B}}K(X_0^{\mathcal{B}}, X_0^{\mathcal{B}}) \end{bmatrix} \\ \mathcal{L}K(X_0, X) &= \begin{bmatrix} \mathcal{A}K(X_0^{\mathcal{A}}, X) \\ \mathcal{B}K(X_0^{\mathcal{B}}, X) \end{bmatrix} \\ \bar{\mathcal{L}}K(X, X_0) &= [\bar{\mathcal{A}}K(X, X_0^{\mathcal{A}}) \quad \bar{\mathcal{B}}K(X, X_0^{\mathcal{B}})] \end{aligned}$$

Here X is to be interpreted as a set of points at which we evaluate those functions drawn from the posterior distribution, in contrast with $X_0 = X_0^{\mathcal{A}} \cup X_0^{\mathcal{B}}$ which is the set of points at which evaluations of the forcing terms are taken.

Proposition 1 (Probabilistic Meshless Method). *Assume \mathcal{A} and \mathcal{B} are linear operators. Then the posterior distribution $\Pi_u^{\mathbf{g}, \mathbf{b}}$ over the solution of the PDE, conditional on the data \mathbf{g}, \mathbf{b} is such that, for $u \sim \Pi_u^{\mathbf{g}, \mathbf{b}}$ we have*

$$\begin{aligned} u(X) &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \\ \boldsymbol{\mu} &= \bar{\mathcal{L}}K(X, X_0) [\mathcal{L}\bar{\mathcal{L}}K(X_0, X_0)]^{-1} [\mathbf{g}^\top \quad \mathbf{b}^\top]^\top \\ \Sigma &= K(X, X) - \bar{\mathcal{L}}K(X, X_0) [\mathcal{L}\bar{\mathcal{L}}K(X_0, X_0)]^{-1} \mathcal{L}K(X_0, X) \end{aligned} \tag{2}$$

Note that the mean function in Eq. 2 is the same as the numerical solution to the PDE that would be obtained using the method of symmetric collocation [5].

Thus far we not discussed the choice of prior covariance k . There are several interesting choices in the literature. Work in [8] proposes use of a covariance which encodes information about the system through its Green's function; [2] examined the properties of this choice in more detail. However, reliance on the Green's function, which is not

in general available in closed-form for complex systems, is a significant drawback. In practice we will generally posit a prior covariance directly by examining the system in question and selecting a prior which encodes a suitable level of differentiability.

We now present a theoretical result describing the rate of convergence of the posterior measure $\Pi_u^{g,b}$. Denote by ρ the differential order of the PDE; that is, the maximum number of derivatives of u required. Furthermore denote by β the smoothness of the prior; the number of weak derivatives that almost surely exist under the prior measure. Lastly, define h to be the “fill distance” of the design points X_0 :

$$h = \sup_{\mathbf{x} \in D} \min_{\mathbf{x}' \in X_0} \|\mathbf{x} - \mathbf{x}'\|_2$$

Theorem 2 (Rate of Convergence). *For a ball $B_\epsilon(u_0)$ of radius ϵ centred on the true solution u_0 of the system (1):*

$$\Pi_u^{g,b}(B_\epsilon(u_0)^c) = \mathcal{O}\left(\frac{h^{2\beta-2\rho-d}}{\epsilon}\right)$$

where c denotes the set complement.

2.1 Illustrative Example: The Forward Problem

We conclude this section by examining the performance of the probabilistic meshless method for a simple 1-dimensional PDE. Consider the system

$$\begin{aligned} -\nabla^2 u(x) &= \sin(2\pi x) & x \in (0, 1) \\ u(x) &= 0 & x = 0, 1 \end{aligned}$$

the solution to which can be computed by direct integration to be $u(x) = -(2\pi^2)^{-2} \sin(2\pi x)$. We compute the PMM solution to this PDE with varying number of design points. In this setting the Green’s function for the system is available explicitly, and so we used its associated prior covariance as suggested in [8]; full details are available in [2].

Samples from the posterior distribution can be seen in Fig. 1; note how, even with 20 design points, there is still significant posterior uncertainty. Convergence plots as the number of design points is increased are shown in Fig. 2

3 Application to Bayesian Inverse Problems

We now turn to an examination of how the PMM, constructed in the previous section, can be applied in Bayesian inverse problems. We now have a system in which we assume the operator \mathcal{A} depends upon some parameter θ , which we emphasise in the below system:

$$\begin{aligned} \mathcal{A}_\theta u(\mathbf{x}) &= g(\mathbf{x}) & \mathbf{x} \in D \\ \mathcal{B}u(\mathbf{x}) &= b(\mathbf{x}) & \mathbf{x} \in \partial D. \end{aligned}$$

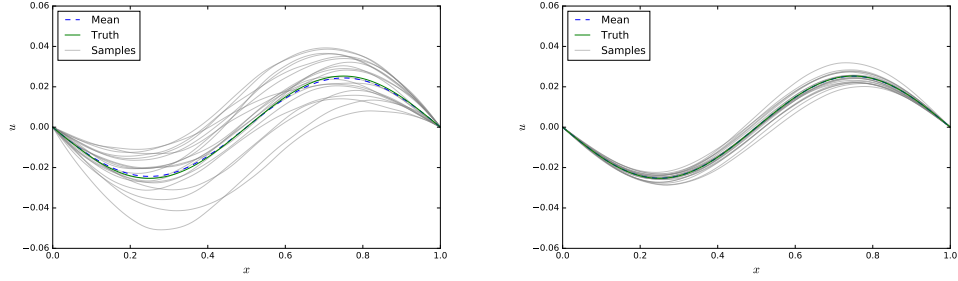


Figure 1: Samples from the posterior distribution over the unknown solution to a one-dimensional PDE, with $m_{\mathcal{A}} = 10$ (left) and $m_{\mathcal{A}} = 40$ (right).

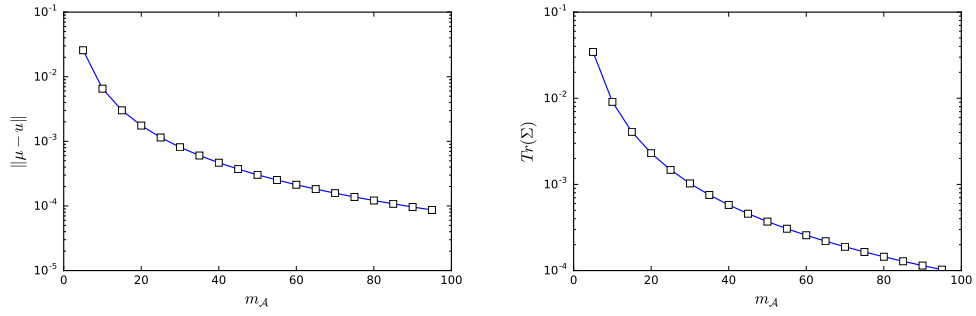


Figure 2: Convergence of mean function (left) and posterior covariance trace (right) as the number of design points $m_{\mathcal{A}}$ is increased.

In a Bayesian inverse problem we place a prior distribution over θ , $\theta \sim \Pi_\theta$, and seek to determine its posterior distribution $\Pi_\theta^\mathbf{y}$ based on data \mathbf{y} collected at locations $\{\mathbf{x}_i\} \subset D$, $i = 1, \dots, n$. Further details on Bayesian inverse problems can be found in [10].

Such a posterior distribution is usually intractable and must be investigated by sampling, which involves solution of the underlying system of PDEs as the sampler visits different values of θ . We assume that the data is obtained by direct observation of the solution u at these locations, corrupted with Gaussian noise

$$y_i = u(\mathbf{x}_i) + \xi_i$$

where $\xi \sim \mathcal{N}(0, \Gamma)$. Our likelihood is thus given by

$$p(\mathbf{y}|\theta, u) = \mathcal{N}(\mathbf{y}; \mathbf{u}, \Gamma) \quad (3)$$

where \mathbf{u}, \mathbf{y} are each vectors in \mathbb{R}^n , with $[\mathbf{u}]_i = u(\mathbf{x}_i; \theta)$ and $[\mathbf{y}]_i = y_i$.

Since the solution u to the PDE system is inaccessible it is common to replace u with an approximation \hat{u} obtained by some numerical scheme. We instead use the PMM as the forward solver, obtaining a measure $\Pi_u^{\mathbf{g}, \mathbf{b}}$ describing our uncertainty. We may then marginalise u in Eq. 3 over this measure to obtain

$$\begin{aligned} p_{\text{PN}}(\mathbf{y}|\theta) &= \int p(\mathbf{y}|\theta, u) \Pi_u^{\mathbf{g}, \mathbf{b}}(du) \\ &= \mathcal{N}(\mathbf{y}; \mu(\theta), \Gamma + \Sigma(\theta)) \end{aligned} \quad (4)$$

where $\mu(\theta), \Sigma(\theta)$ are as in Prop. 1, and we have emphasised the dependence on θ . This is thus similar to the standard approach of replacing u with \hat{u} in Eq. 3, but we compensate for the inaccuracy of the forward solver with an additive covariance term Σ incorporating the uncertainty in the posterior distribution for the forward problem.

We now present a result which guarantees consistency in the inverse problem when we replace the likelihood in Eq. 3 with that in Eq. 4.

Proposition 3. (*Inverse Problem Consistency*) Let $\Pi_{\theta, \text{PN}}^\mathbf{y}$ be the posterior distribution which uses the PN likelihood given in Eq. 4. Assume that the posterior distribution $\Pi_\theta^\mathbf{y}$ contracts such that $\Pi_\theta^\mathbf{y} \rightarrow \delta(\theta_0)$ as $n \rightarrow \infty$, a Dirac measure centred on the true value of θ , θ_0 . Then $\Pi_{\theta, \text{PN}}^\mathbf{y}$ contracts such that $\Pi_{\theta, \text{PN}}^\mathbf{y} \rightarrow \delta(\theta_0)$ provided

$$h = o(n^{-1/(\beta - \rho - d/2)})$$

3.1 Illustrative Example: The Inverse Problem

We now return to the previous illustrative example to demonstrate the use of a probabilistic solver in the inverse problem. Consider the system

$$\begin{aligned} -\nabla \cdot \theta \nabla u(x) &= \sin(2\pi x) & x \in (0, 1) \\ u(x) &= 0 & x = 0, 1 \end{aligned}$$

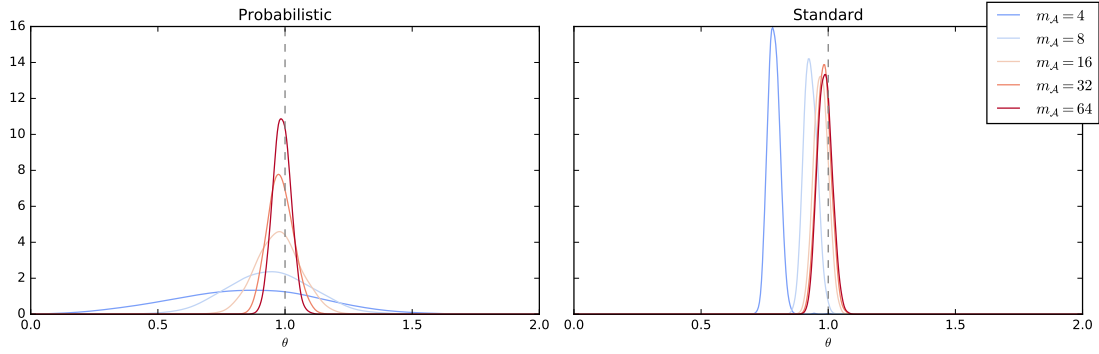


Figure 3: Posterior distributions over θ with varying numbers of design points, on the left using the PMM, and on the right the standard approach of using a plug-in estimate for the PDE solution, here given by symmetric collocation.

with the goal of inferring the parameter θ . Data y_i was generated from the explicit solution to this problem with $\theta = 1$ at locations $x = 0.25, 0.75$, and corrupted with Gaussian noise with distribution $\mathcal{N}(0, 0.01^2)$.

In Fig. 3 we compare posterior distributions for θ generated with the PMM versus the standard approach of plugging a numerical solution of the PDE into the likelihood and ignoring discretisation error. The numerical method used in the standard approach was symmetric collocation, the most natural comparison. Note that when using collocation the posterior distributions are peaked and biased for small m_A , and that the posterior uncertainty does not appear to depend on the number of design points. Conversely when using the probabilistic method we see that for small m_A the posterior distributions are wide and flat, while as m_A increases the distributions peak and centre on the true value of θ . Thus, with a standard numerical method the posteriors over θ do not take into account the quality of the numerical solver used; for poor forward solvers based on coarse discretisations, the posteriors produced are as confident as those produced with a fine, accurate numerical solver. With a probabilistic forward solver the variance in the forward solver is propagated into the inverse problem, resulting in robust inferences even when the discretisation is coarse.

4 A Nonlinear Example

We now present an application of the methods discussed herein to a nonlinear partial differential equation known as the steady-state Allen–Cahn system, a model from mathematical physics describing the motion of boundaries between phases in iron alloys. This

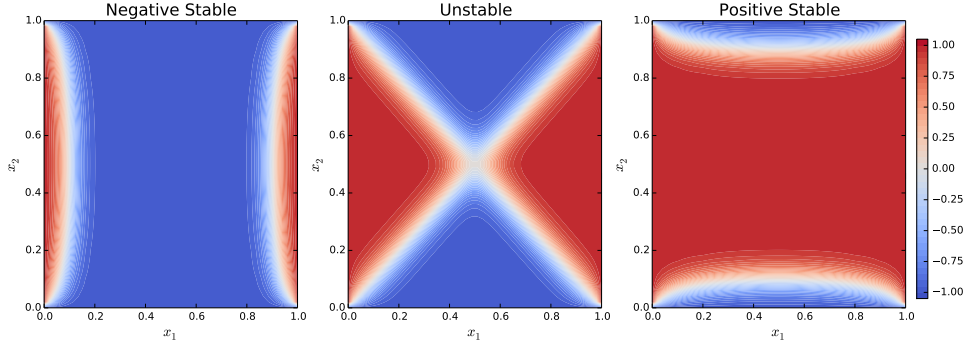


Figure 4: Solutions to the Allen–Cahn system for $\delta = 0.04$

is given by

$$\begin{aligned}
 -\delta \nabla^2 u + \delta^{-1}(u^3 - u) &= 0 & \mathbf{x} \in (0, 1)^2 \\
 u &= +1 & x_1 \in \{0, 1\}, x_2 \in (0, 1) \\
 u &= -1 & x_2 \in \{0, 1\}, x_1 \in (0, 1)
 \end{aligned} \tag{5}$$

We phrase this as an inverse problem for determining δ . This system is noteworthy for the fact that it does not admit a unique solution; the three solutions to this system for $\delta = 0.04$ are shown in Fig. 4. These were generated using the deflation technique described in [4].

Since this is a nonlinear system the posterior distribution will not be Gaussian, and we must resort to sampling techniques to explore the posterior distribution. In brief, we introduce a latent function z and rearrange the system as follows:

$$-\delta \nabla^2 u - \delta^{-1} u = z \tag{6}$$

$$\delta^{-1} u^3 = -z \tag{7}$$

Note that by adding Eq. 6 and Eq. 7 we return to the original equation describing the interior dynamics given in Eq. 5. However Eq. 7 is monotonic and thus invertible; by inverting this we arrive at a new system:

$$\begin{aligned}
 -\delta \nabla^2 u - \delta^{-1} u &= z \\
 u &= (-\delta z)^{1/3}
 \end{aligned}$$

This system is equivalent to the original system but, importantly, is *linear*. Thus by the introduction of z we are able to arrive at a new system which can be solved using the PMM.

It remains to describe z , a latent function whose value is unknown. We seek to marginalise z in the likelihood

$$p(\mathbf{y}|\delta) = \int p(z|\delta) \int p(\mathbf{y}|u) \Pi_u^{\mathbf{g}, \mathbf{b}, z}(du) dz \tag{8}$$

where $\Pi_u^{g,b,z}$ is now additionally conditioned on a known value for z . This integral is intractable. However by sampling from the posterior distribution over δ by pseudo-marginal MCMC it is sufficient to produce an unbiased estimate of this quantity. This is accomplished by importance sampling; we assume an improper prior $p(z|\delta) \propto 1$ and approximate Eq. 8 by the Monte-Carlo estimate

$$p(\mathbf{y}|\delta) \approx \frac{1}{M} \sum_{i=1}^M \frac{\int p(\mathbf{y}|u) \Pi_u^{g,b,z_i}(du)}{r(z_i|\mathbf{y}, \delta)}$$

for $z_i \sim r(z|\mathbf{y}, \delta)$.

The importance distribution $r(z|\mathbf{y}, \delta)$ is chosen by solving the original system in Eq. 5 using the techniques described in [4], with a coarse finite-element solver. This gives estimates $\{\hat{u}_1, \hat{u}_2, \hat{u}_3\}$ for the solution given a value of δ . By applying Eq. 7 to these estimates we obtain estimates of three values of z ; $\{\hat{z}_1, \hat{z}_2, \hat{z}_3\}$.

To handle the multimodality in the solutions we extend the state-space of the inverse problem to include the solution index j . The importance distribution is constructed as a Gaussian distribution

$$z \sim \mathcal{GP}(\hat{z}_j, k)$$

with $r(z|\mathbf{y}, \delta, j)$ thus the appropriate multivariate Gaussian density after the field for z has been discretised. Discretisation points are necessarily chosen to match X_0^A , the design points for u in the interior of the domain.

For application of the PMM we choose a squared-exponential prior covariance

$$k(x, x') = \exp \left(-\frac{\|x - x'\|_2^2}{2\ell^2} \right)$$

which is known to describe infinitely-differentiable functions. This choice is motivated by the high differential order required by the PDE; since we must be able to apply both the operator and the adjoint to the kernel, in this case we require that the covariance be twice differentiable in each argument, which amounts to a four-times differentiable covariance if the covariance chosen is isotropic.

The length-scale hyper-parameter ℓ was incorporated into the MCMC procedure, endowed with a half-Cauchy hyper-prior as recommended in [6]. The parameter of interest δ was endowed with a uniform prior over the interval $(0.02, 0.15)$, in which the PDE was empirically found to consistently have three solutions.

Posterior distributions for δ generated using this methodology are shown in Fig. 5; these are compared with posterior distributions generated using a finite-element forward solver. In the finite-element case we see a more extreme version of the bias shown in Fig. 3 for coarse grids, whereas when using a probabilistic forward solver the posteriors are once again wider to account for an inaccurate forward solver.

We should also comment on the comparison to the finite-element method here; in the previous example comparison was to the symmetric collocation method for solving PDEs; in this case the comparison is more direct as the solution for the PDE in symmetric collocation is simply the posterior mean from the PMM. In this case we use a

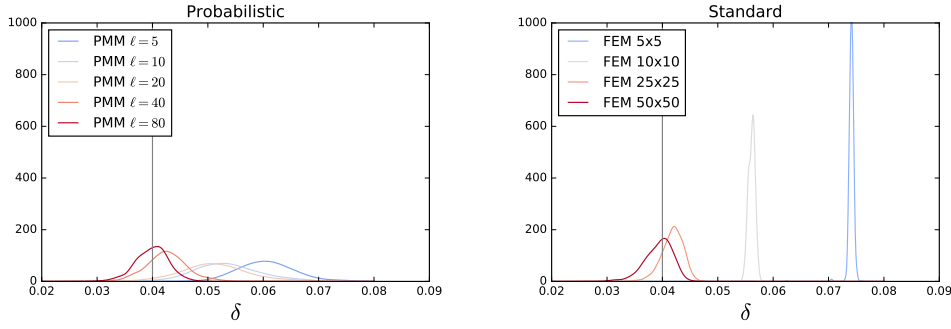


Figure 5: Posterior distributions for δ obtained by use of the technique described herein (left) versus a standard Finite Element solver that does not model discretisation error (right).

finite-element solver both to highlight the fact that the behaviour witnessed when using symmetric collocation is not unique to that solver, and because in existing methods for finding the multiple solutions to the Allen-Cahn equation the base numerical method applied is the finite-element method. Furthermore we note that as the underlying numerical method becomes arbitrarily accurate, the posterior inferences made in the inverse problem should be invariant to the forward solver used.

5 Discussion

We have shown how to construct probabilistic models for the solution of partial differential equations, which quantify the uncertainty arising from numerical discretisation of the system. We have further shown how the uncertainty in the forward problem can be propagated into posteriors over parameters in inverse problems. This allows robust inferences to be made in inverse problems, even when the numerical scheme used to solve the forward problem is inaccurate, which is useful in cases where obtaining highly accurate solutions is computationally expensive, or where we are willing to tolerate less certain inferences in exchange for fast computation. In particular we have illustrated how this might be used to make inferences in nonlinear systems where a variety of phenomena, such as a non-unique solution could cause a numerical solver to fail.

Immediate extensions to this work lie in examining evolutionary systems in which the solution is additionally a function of time; the added complexity from the additional dimension demands more focussed attention. We also seek to examine a more generic approach for sampling from posterior distributions for nonlinear PDEs. Furthermore we note that the observations we have chosen for the forward problem are only one possible choice; another attractive option is given by Galerkin schemes for approximating PDEs, by choosing our observations to be Galerkin projections.

Lastly we seek to explore other choices of prior. The Gaussian measure is an unrealistic option in general, as it penalises extreme values and prevents encoding such simple properties as positivity of solutions.

6 Acknowledgements

TJS was supported by the Free University of Berlin within the Excellence Initiative of the German Research Foundation (DFG). MG was supported by EPSRC [EP/J016934/1, EP/K034154/1], an EPSRC Established Career Fellowship, the EU grant [EU/259348] and a Royal Society Wolfson Research Merit Award.

The authors would like to thank John Skilling for useful discussion, Patrick Farrell for providing code used in generating these results and François-Xavier Briol for helpful feedback. In addition they express gratitude to the developers of the Python libraries Autograd and GPyOpt.

References

- [1] Igor Cialenco, Gregory E Fasshauer, and Qi Ye. Approximation of stochastic partial differential equations by a kernel-based collocation method. *International Journal of Computer Mathematics*, 89(18):2543–2561, December 2012.
- [2] Jon Cockayne, Chris Oates, Tim Sullivan, and Mark Girolami. Probabilistic Meshless Methods for Partial Differential Equations and Bayesian Inverse Problems. *arXiv:1605.07811v1*, May 2016.
- [3] Patrick R Conrad, Mark Girolami, Simo Särkkä, Andrew Stuart, and Konstantinos Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Statistics and Computing*, 2016.
- [4] Patrick E Farrell, Asgeir Birkisson, and Simon W Funke. Deflation techniques for finding distinct solutions of nonlinear partial differential equations. *SIAM Journal on Scientific Computing*, 37(4):A2026–A2045, 2015.
- [5] Gregory E Fasshauer. Solving differential equations with radial basis functions: multilevel methods and smoothing. *Advances in Computational Mathematics*, 11(2-3):139–159, 1999.
- [6] A Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3):515–534, 2006.
- [7] Philipp Hennig, Michael A Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. *Proc R Soc A*, 471(2179):20150142, July 2015.
- [8] Houman Owhadi. Bayesian numerical homogenization. *Multiscale Modeling & Simulation*, 13(3):812–828, 2015.
- [9] Houman Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from Hierarchical Information Games. *arXiv:1503.03467v4*, March 2015.

- [10] Andrew M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19: 451–559, 2010. ISSN 0962-4929.
- [11] Henryk Woźniakowski. What is information-based complexity? *Essays on the complexity of continuous problems*, pages 89–95, 2009.